

УДК 004.42

ОБЗОР РЕАЛИЗАЦИЙ GPGPU. ВВЕДЕНИЕ В ТЕХНОЛОГИЮ NVIDIA CUDA

Ефремов А.А.

Научный руководитель – к.ф.-м.н., доцент Карпова Е.Д.

Сибирский федеральный университет

Работа посвящена обсуждению техники программирования GPGPU, а также основных реализаций данной техники. Особое внимание уделяется представленной компанией nVidia технологии CUDA. Приведены примеры программ написанных с помощью технологии CUDA.

GPGPU (англ. General-purpose graphics processing units — «графический процессор общего назначения») — техника использования графического процессора видеокарты для общих вычислений.

Среди известных в настоящее время технологий для GPGPU можно отметить следующие наиболее популярные инструменты.

- OpenCL – фреймворк (программное обеспечение, облегчающее разработку и объединение разных компонентов большого программного проекта) для написания компьютерных программ, связанных с параллельными вычислениями на различных графических и центральных процессорах.
- ATI Stream Technology - это набор аппаратных и программных технологий, которые позволяют использовать графические процессоры AMD, совместно с центральным процессором, для ускорения многих приложений (не только графических).
- DirectCompute - интерфейс программирования приложений, который входит в состав DirectX - набора API от Microsoft.
- CUDA - (Compute Unified Device Architecture) — программно-аппаратная архитектура, разработка компании nVidia. На сегодняшний день основная технология программирования GPGPU.

Опишем более подробно самую популярную на сегодняшний момент архитектуру nVidia CUDA.

CUDA (англ. Compute Unified Device Architecture) — программно-аппаратная архитектура, позволяющая производить вычисления с использованием графических процессоров NVIDIA, поддерживающих технологию GPGPU. В работе обсуждаются следующие вопросы, касающиеся этой технологии.

1. *Основные характеристики CUDA.* Приводится краткое описание технологии CUDA. Сравниваются архитектуры CPU и GPU.
2. *Описание аппаратной реализации CUDA.* Описаны достоинства данной технологии относительно решений других компаний. Исследованы архитектурные недостатки аппаратной части CUDA.
3. *Описание программной реализации CUDA.* Состав программной части технологии. Краткое описание инструментов, предоставляемых компанией nVidia.
4. *Пример использования CUDA:* перемножение матриц. Автором написаны две программы, реализующие перемножение матриц. Реализован последовательный алгоритм решения на CPU на языке C/C++, а также параллельный алгоритм на языке CUDA C/C++. Для расчетов использовался процессор Intel Core 2 Duo E4600. Для расчетов на CUDA – nVidia GeForce 250 GTS.

5. *Некоторые особенности при разработке программ на CUDA Си.* Приведен пример несовершенства существующей программной реализации CUDA на функции AtomicMax (AtomicMin). Предложен метод, который поможет обойти ограничения данной функции.

Реализация технологии GPGPU CUDA от компании nVidia на сегодняшний день является фаворитом в данной области. Технологии конкурентов (ATI Stream и др.) пока не имеют такой развитой аппаратной и программной реализации, как у nVidia. Постоянный технический прогресс в развитии графических чипов общего назначения, а также стремительное развитие nVidia CUDA SDK позволяют считать компанию nVidia лидером в области высокопроизводительных вычислений с использованием GPGPU.